Using Standardized Encounters to Understand Reported Racial/Ethnic Disparities in Patient Experiences with Care

Robin M. Weinick, Marc N. Elliott, Angelo E. Volandes, Lenny Lopez, Q Burkhart, and Mark Schlesinger

Objective. To assess the extent to which racial/ethnic differences in ratings of patient experiences with health care represent true differences versus differences in expectations, how scales are used, or how identical physician–patient interactions are perceived by members of different groups.

Study Setting. Primary data collection from a nationally representative online panel ($n = 567$), including white, African American, and Latino respondents.

Study Design. We administered questions on expectations of care, a series of written vignettes, a video-depicted doctor–patient interaction, and modified CAHPS Clinician and Group Doctor Communication items.

Principal Findings. Different groups reported generally similar expectations regarding physicians' behaviors and provided similar mean responses to CAHPS communication items in response to standardized encounters.

Conclusions. Preliminary evidence suggests that unlike more subjective global ratings, reported disparities in more specific and objective CAHPS composites may primarily reflect differences in experiences, rather than differences in expectations and scale use, adding to our confidence in using the latter to assess disparities.

Key Words. Racial/ethnic disparities, CAHPS, patient experiences with care, measurement

A growing body of research demonstrates that patients from different racial and ethnic groups report differing experiences with the health care system when using well-validated measurement tools such as CAHPS. However, there are seeming paradoxes within these observed differences. For example, African American patients provide higher global ratings of their care and personal clinicians than white patients, despite more objective reports that

their experiences are problematic, including having worse communication and less responsive providers (Morales et al. 2001; Lurie et al. 2003; Uhrig et al. 2004; Dayton et al. 2006). Similarly, Latinos are more likely than whites to report problems getting needed care and with respect, but they simultaneously provide higher global ratings of their doctors (Weech-Maldonado et al. 2003, 2004, 2008).

Understanding these seemingly paradoxical results could help in describing the role that race and ethnicity play in mediating clinician–patient interactions. There is evidence that this paradox may be partially explained by differences in scale use for global ratings, such as extreme response tendency (ERT) (Elliott et al. 2009b), that do not apply to more specific, objective "report" items. This lack of comparability may make global ratings less suitable for assessing and reporting racial/ethnic disparities, despite the appeal of a single measure that is easily interpreted by the public. This issue is of more than academic interest, as the results of patient experiences with care surveys are increasingly disseminated to prospective patients. For example, the 2008 expansion of Hospital Compare makes such data from the CAHPS Hospital Survey publicly available for individual hospitals (http://www.hospitalcompare.hhs.gov), and the Medicare Improvements for Patients and Providers Action of 2008 mandates public reporting of patient experiences with Medicare plans by race/ethnicity.

Considering race/ethnicity exacerbates challenges to assessing complex experiences such as clinical encounters, because the construct embodies a lifelong set of experiences that can fundamentally transform how medical episodes are perceived, evaluated, or described. As a result, one potential explanation for African Americans' and Latinos' higher scores on global rating scales despite also reporting more frequent problems is that they have systematically lower expectations of medical care and thus may be more easily satisfied (Weech-Maldonado et al. 2008), in keeping with evidence that expectations affect patients' evaluations of care (Jackson, Chamberlin, and Kroenke 2001; Noble et al. 2006). Alternatively, various racial/ethnic groups may use rating scales differently when responding to questions on experiences

with care. Previous studies have found that Latino and African American respondents use health care response scales differently than white respondents (Gallagher, Fowler, and Cleary 2004; Weech-Maldonado et al. 2008; Elliott et al. 2009b). Finally, members of different racial/ethnic groups may have systematically varying interpretations of identical interactions because they value particular aspects of these interactions in different ways.

Robust, comparable measurement of patient experiences is essential to designing effective interventions to reduce well-documented racial and ethnic disparities in health care; little progress can be made if the roles of scale use, expectations, preferences, and experiences are not clearly understood. This study was designed to assess the extent to which African American, Latino, and white respondents provide similar responses to items from the CAHPS Clinician and Group Survey in response to standardized clinical scenarios in order to clarify the interpretation of racial/ethnic differences in patient-reported real-world health care experiences. We rely in part on methodology originally developed by King et al. (2004) to assess the extent of cross-cultural incomparability in survey responses by presenting an ordered series of short vignettes and examining the extent to which different groups of individuals vary in the responses they offer to questions about the vignettes.

METHODS

Sample

This study was conducted using the Knowledge Networks panel, an ongoing Internet panel based on a random digit dialing sample of the full U.S. adult population, which is designed to be nationally representative and has a 54.9 percent participation rate (Knowledge Networks No Date a). The panel provides free Web TV access for those who do not have a home Internet connection, representing lower Income adults who would otherwise be disproportionately excluded. A variety of health-related studies have used this panel and support its validity (Schlenger et al. 2002; Silver et al. 2002; Baker et al. 2003; Wagner et al. 2004). Each individual is contacted by e-mail regarding participation in an individual survey effort. Individuals who do not complete the survey within the requested timeframe are sent e-mail reminders; if this does not generate a response, the individual receives a reminder phone call.

A random sample of 1,275 English-speaking adults was selected from the panel, stratified by race/ethnicity to obtain similar numbers of respondents from each of three racial/ethnic groups, and resulted in 567 responses (44.5

percent completion rate within the Knowledge Networks panel). Respondents included 204 white (completion rate 52.3 percent), 163 black (completion rate 36.9 percent), and 200 Latino (completion rate 45.0 percent) adults. Asians were excluded from the sample, because the panel included too few Asian members to allow for precise estimation. The study was fielded in October and November 2008.

Study Procedures

Respondents answered questions regarding expectations, assessments of written vignettes, and assessments of a videotaped simulated physician–patient encounter. The written vignettes allowed us to efficiently expose respondents to multiple scenarios depicting a gradation of physician responsiveness to patients, which provides insight into differential use of response scales. In contrast, the video provides greater realism, and previous research has shown that people respond differently to video descriptions of health care concerns than to equivalent narrative descriptions (Volandes et al. 2007). Consequently, the video supported detailed measurement and comparison of perceptions of positive and negative aspects of physician behavior.

In the first part of the study, we asked respondents to answer a series of questions that have previously been used to assess expectations about physicians' behavior in general (Gary 2006). Here, we use the term "expectations" as a shorthand for a complex construct that is based on beliefs and past experiences. These expectations have likely evolved over time for each respondent based on his/her beliefs and previous experiences with physicians. These questions asked: "Roughly how many doctors do you think:

- Take the time and effort to learn about the most up-to-date treatments and drugs?
- Don't take enough time to talk with patients about their medical care?
- Will speak up for their patients in disputes with health insurance plans?
- Make too many mistakes in taking care of their patients?
- Treat all patients fairly regardless of race?

The response options were *no doctors at all, some doctors, most doctors,* and *all doctors.*

Next, each respondent reads a series of five vignettes describing interactions between patients and physicians. Each vignette began with an identical

description of a patient complaining of headaches. The vignettes differed in physician responsiveness to the patient's concerns, and they were reviewed by multiple team members to ensure that they depicted differential levels of physician responsiveness to the patient's concerns and avoided any appearance of the physician reacting to an escalating clinical situation. This approach follows King et al.'s (2004) technique of exposing respondents to multiple written vignettes that are constructed to differ in terms of a well-defined characteristic (here, physician responsiveness) and testing the extent to which respondents differentiate among those vignettes on a fixed response scale (here, modified items from the CAHPS Clinician and Group survey). The five vignettes appear in Appendix SA2. Although the vignettes were presented to respondents in randomized order, we refer to them here as Vignettes 1 (least responsive) through 5 (most responsive). The vignettes thus constitute an ordinally scaled measure of physician responsiveness. Vignettes were constructed so their length was independent of the degree of physician responsiveness.

After reading each vignette, respondents were asked modified versions of three of the six items within the Doctor Communication composite of the CAHPS Clinician and Group Survey, modified for this setting:

- To what extent did this doctor listen carefully to [the patient]?
- To what extent did this doctor show respect for what [the patient] had to say?
- To what extent did this doctor spend enough time talking to [the patient] about his headaches?

The CAHPS questions were originally designed to ask how frequently the patient's physician exhibited each of the behaviors across what might have been multiple real-life encounters, with response options of *never*, *sometimes*, *usually*, or *always*. Because this format is not consistent with a single encounter described in a vignette, we modified the stem and response scale to ask about the extent to which the doctor exhibited each of these behaviors (*not at all*, *very little*, *to some extent*, or *to a great extent*). To minimize respondent burden, the three questions from this composite deemed less relevant to the vignettes were omitted from this study.

Finally, we prepared a 4-minute video simulating a single physician–patient encounter, in which a diabetic patient's lack of success at controlling her blood sugar is discussed. The video makes clear that the doctor and patient have a longstanding, comfortable relationship. The physician indicates how

busy he is and responds to the patient's lack of progress with significant frustration, though he also tries to encourage her and discusses alternative improvement strategies. A transcript of the video is included in Appendix SA3. The script and final video were reviewed for accuracy and tone by two physicians (A. E. V. and L. L.) and the principal investigator (R. M. W.), with the overall tone intended to roughly balance positive and negative physician behaviors.

Following the video viewing, each respondent was asked to provide a 0–10 global rating of the doctor and to answer a series of five questions from the Doctor Communication composite of the CAHPS Clinician and Group Survey, modified as described for the written vignettes. The questions are listed in Table 3. The video was followed by more extensive follow-up than the written vignettes for two reasons. First, the written vignettes involved very short descriptions, making it unlikely that respondents had adequate information on which to base a response to the global rating question or to a more extensive set of follow-up questions. Second, to avoid undue respondent burden, we did not ask additional questions multiple times for each of the five vignettes.

Finally, each respondent in the full sample was asked a series of questions to elucidate the rationale for his or her response to the video. These questions applied an approach drawn from Motivational Interviewing (Miller and Rollnick 1991; Burke, Arkowitz, and Menchola 2003), asking respondents why they provided the responses they did to the CAHPS global rating question, rather than higher or lower responses. This technique helps elucidate individuals' motivations, and here it helps to clarify respondents' perceptions of the physician in the video and salient aspects of his behavior that influenced their rating. These questions included a set of 10 positive and 10 negative characteristics, with responses on a four-level ordinal scale, assessing the extent to which each characteristic described the physician (*not at all*, *very little*, *to some extent*, or *to a great extent*). The positive characteristics included the extent to which the doctor was knowledgeable, motivated the patient, had a good relationship with the patient, was reassuring, was trusted by the patient, was kind, was respectful, was understanding, liked the patient, and was helpful; the negative characteristics were the extent to which the doctor was rushed, was arrogant, was impatient, was pushy, ignored the patient's questions, was disrespectful, was intimidating, was disapproving, interrupted the patient, and disliked the patient. More detail on the development of these questions is shown in Appendix SA4.

This study was reviewed by the Partners HealthCare Human Research Committee (Boston, MA).

Analysis

Our analyses included both bivariate and multivariate statistical models.

For each of the five measures of expectations, means were computed by race/ethnicity, and African Americans and Latinos were compared with whites via independent sample *t*-tests.

For each of the three modified CAHPS communication items and each of the five vignettes, mean responses were calculated by race/ethnicity. A series of three multivariate linear regressions predicted responses to each CAHPS item from indicators of physician responsiveness, case-mix adjustors (age, gender, education), indicators of African American and Latino race/ethnicity, and the interaction between physician responsiveness and race/ethnicity. Additional models parameterized physician responsiveness linearly, rather than as categories, for greater power to detect disparities. These multivariate models adjusted for the correlation of responses to multiple vignettes for each respondent using the Huber–White sandwich estimator of variance.

For each of the five modified CAHPS items assessed for the video, as well as for the indices of perceived positive and negative physician behavior, case-mix-adjusted mean responses to the video were calculated and compared by race/ethnicity. The indices of perceived positive and negative behavior were constructed via an exploratory factor analysis of responses to the 20 attributes described above from the full online sample ($n = 567$). This analysis identified two factors (regardless of whether Pearson correlations or polychoric correlations were used), resulting in one positive and one negative index of perceived physician behavior, each constructed as the mean of 10 items. As described above, these two indices help to elucidate respondents' rationale for their global rating of the physician in the video.

Similar analyses were performed for the 0–10 global rating of the physician in the video. In addition, ERT by race/ethnicity was assessed in two ways. First, we compared the standard deviation of the responses by race/ethnicity using the Levene test. Second, multinomial logistic regression was used to test the four most extreme responses (0–1 pooled; 9–10 pooled) relative to the middle seven responses (2–8 pooled), similar to the approach used by Weech-Maldonado et al. (2008) and Elliott et al. (2009b). Finally, a multivariate model predicted the 0–10 global rating of the physician in the video from indicators of African American and Latino race/ethnicity, case-mix adjustors, indices of perceived positive and negative physician behavior, and the interaction of race/ethnicity with perceived physician behavior. The

interaction helped assess the extent to which perceptions of physician behavior differentially affect global ratings across racial/ethnic groups.

Statistical analyses were conducted using *STATA 10* (Stata Corp, College Station, TX, USA) and *SAS 9.2* (SAS Institute Inc., Cary, NC, USA). Unless otherwise noted, $p < .05$ using two-sided tests for any differences discussed in the text.

RESULTS

Table 1 displays the demographic characteristics of our sample, overall and by racial/ethnic group. By design, the racial/ethnic distribution of our sample differs considerably from that of the U.S. population. Our sample is slightly older than the U.S. population as a whole and the full Knowledge Networks panel, and more likely to be in the middle income group than the overall U.S. population (DeNavas-Walt, Proctor, and Smith 2008; U.S. Census Bureau 2009a; Knowledge Networks No Date b). At the same time, the distribution of our sample by gender, education, and area of residence looks similar to both the broader U.S. population and the full Knowledge Networks panel (U.S. Census Bureau 2009a, b; U.S. Census Bureau No Date a; U.S. Census Bureau No Date b). Racial/ethnic differences in these characteristics are similar to those in the general population.

Table 2 shows mean expectations regarding physician behavior by race/ethnicity. Average responses tend to fall near the middle of the scale ("some" to "most" doctors), regardless of respondent race/ethnicity. Mean responses range from an overall average of 2.21–2.78 for positive behaviors, and 2.06–2.40 for negative behaviors, where 1 corresponds to "no doctors at all" and 4 to "all doctors." Notably, the only expectation for which there were statistically significant differences by race/ethnicity relates to whether physicians treat all patients fairly regardless of their race, with both African American (mean 2.53) and Latino (mean 2.78) respondents believing that fewer doctors do so than white respondents (mean 2.98), equivalent to 45 percent of African Americans and 20 percent of Latinos shifting responses from one category to an adjacent, more positive category. The remaining behaviors demonstrate no statistically significant racial/ethnic differences ($p > .05$ in all cases).

Table 3 summarizes CAHPS responses to the vignettes (panel A) and the video of the simulated encounter (panel B). For each of the three CAHPS items, panel A shows the case-mix-adjusted mean response to each of the five written vignettes by race/ethnicity. Responses are increasingly positive with increasing

Table 1:   Demographic Characteristics of Respondents by Race/Ethnicity

	Percent (SE)			
	All (n = 567)	White (n = 204)	African American (n = 163)	Latino (n = 200)
Race/ethnicity				
White	36 (2)			
African American	29 (2)			
Latino	35 (2)			
Age				
18–44	44 (2)	41 (3)	37 (4)	54 (4)**
45–64	40 (2)	40 (3)	50 (4)	33 (3)
65+	15 (2)	20 (3)	13 (3)	13 (2)
Gender				
Male	44 (2)	47 (3)	40 (4)	44 (4)
Female	56(2)	53 (3)	60 (4)	56 (4)
Education				
Less than high school	12 (1)	8 (2)	12 (3)	17 (3)**
High school graduate	28 (2)	32 (3)	29 (4)	23 (3)*
Some college	34 (2)	31 (3)	34 (4)	38 (3)
Bachelors degree or higher	26 (2)	29 (3)	26 (3)	23 (3)
Household income				
<$35,000	26 (2)	20 (3)	35 (4)**	27 (3)
$35,000–99,999	60 (2)	63 (3)	58 (4)	59 (3)
$100,000+	14 (1)	18 (3)	7 (2)**	15 (3)
Lives in Metropolitan Statistical Area				
Yes	91(1)	83 (3)	96 (2)***	96 (1)***
No	9 (1)	17 (3)	4 (2)***	4 (1)***
Region of country				
Northeast	15 (2)	15 (2)	16 (3)	16 (3)
Midwest	18 (2)	22 (3)	19 (3)	13 (2)*
South	41 (2)	40 (3)	52 (4)*	32 (3)
West	26 (2)	24 (3)	13 (3)*	40 (3)***

*$p<.05$,
**$p<.01$,
***$p<.001$ for test of characteristic mean in the designated racial/ethnic group differing from the mean among whites.

depicted physician responsiveness to the patient. These findings are replicated in additional linear regressions (not shown), which found significant positive coefficients for physician responsiveness ($\beta = 0.56$, 0.63, and 0.60 points per level of linearly coded responsiveness for *listen*, *respect*, and *time*, respectively; $p<.001$ for each), confirming that the written vignettes effectively conveyed the intended systematically increasing degree of physician responsiveness.

Table 2:   Expectations Regarding Physician Behavior by Race/Ethnicity

| | *Race/Ethnicity* | | | | | | | |
| | *All* | | *White* | | *African American* | | *Latino* | |
Roughly how many doctors do you think:	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*
Positive behaviors								
Take the time and effort to learn about the most up-to-date treatments and drugs?	2.56	0.03	2.63	0.04	2.52	0.05	2.52	0.05
Will speak up for their patients in disputes with health insurance plans?	2.21	0.03	2.22	0.05	2.25	0.05	2.18	0.05
Treat all patients fairly regardless of race?	2.78	0.03	2.98	0.04	2.53*	0.05	2.78*	0.05
Negative behaviors								
Don't take enough time to talk with patients about their medical care?	2.40	0.03	2.44	0.04	2.33	0.05	2.42	0.05
Make too many mistakes in taking care of their patients?	2.06	0.02	2.09	0.03	2.03	0.03	2.05	0.04

Response options are 1 = no doctors at all; 2 = some doctors; 3 = most doctors; 4 = all doctors.
*$p < .05$ versus white.

At the same time, however, responses are quite similar by race/ethnicity within a given vignette ($p > .05$ in all instances). Case-mix-adjusted repeated-measures multivariate models that were designed to maximize statistical power to detect racial/ethnic differences confirmed this (not shown). In these same models, there was no significant association of African American or Latino race/ethnicity, as compared with white, with CAHPS responses ($p > .4$ in all instances), suggesting that the three groups responded to the CAHPS reports items and used the response scales in similar ways when presented with identical written stimulus material. Given the absence of evident disparities here, additional analytic techniques applicable to these vignettes (King et al. 2004) were not pursued.

Panel B of Table 3 shows mean responses to the CAHPS questions that were asked of each respondent following the video viewing. Notably, the mean 0–10 rating of the doctor was below 5 for all racial/ethnic groups, suggesting that the physician in this third-person encounter was perceived more negatively than is typical of perceptions of one's own physician in the real world, given that means near 9 are more typical for such ratings of one's own physician (e.g., Elliott et al. 2009b). Coefficients in case-mix-adjusted regressions showed no evidence of racial/ethnic differences in responses to any of the five modified Doctor Communication items or in responses to the global rating. A repeated-measures multiple regression (similar to the model

Table 3:   Case-Mix-Adjusted Mean CAHPS Responses to Written Vignettes and Video Interaction by Race/Ethnicity (Adjusted for Age, Education, and Gender)

	(A) Written Vignettes							
	*Race/Ethnicity**							
To what extent did this doctor:	*All*		*White*		*African American*		*Latino*	
	Mean	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*
. . . listen carefully to the patient?[†]								
Vignette/physician responsiveness[‡]								
1	1.66	0.03	1.64	0.05	1.62	0.06	1.71	0.06
2	1.90	0.03	1.91	0.06	1.82	0.07	1.94	0.06
3	2.79	0.03	2.75	0.05	2.76	0.06	2.84	0.05
4	3.27	0.03	3.30	0.05	3.28	0.06	3.22	0.06
5	3.65	0.03	3.74	0.04	3.64	0.05	3.57	0.05
All	2.65	0.02	2.67	0.03	2.54	0.07	2.66	0.03
. . . show respect for what the patient had to say?[†]								
Vignette/physician responsiveness[‡]								
1	1.43	0.03	1.41	0.05	1.44	0.06	1.43	0.05
2	1.73	0.03	1.73	0.05	1.71	0.07	1.73	0.06
3	2.72	0.04	2.69	0.06	2.71	0.07	2.75	0.06
4	3.27	0.03	3.30	0.06	3.27	0.07	3.24	0.04
5	3.64	0.03	3.76	0.04	3.62	0.05	3.55	0.05
All	2.56	0.02	2.58	0.03	2.49	0.07	2.54	0.03
. . . spend enough time talking to the patient about his headaches?[†]								
Vignette/physician responsiveness[‡]								
1	1.36	0.03	1.35	0.04	1.34	0.05	1.39	0.04
2	1.57	0.03	1.56	0.05	1.51	0.07	1.60	0.05
3	2.33	0.04	2.26	0.06	2.25	0.08	2.44	0.06
4	2.98	0.04	3.01	0.07	2.92	0.08	2.99	0.07
5	3.56	0.03	3.65	0.05	3.54	0.06	3.49	0.06
All	2.36	0.02	2.37	0.03	2.28	0.07	2.38	0.03

	(B) Video Interaction							
CAHPS Communication Composite Items[†]								
To what extent did this doctor:								
Listen carefully?			2.43	0.04	2.33	0.07	2.46 0.07	2.50 0.07
Show respect for what the patient had to say?			2.18	0.04	2.08	0.07	2.26 0.07	2.22 0.07
Spend enough time with the patient?			2.64	0.04	2.61	0.07	2.66 0.07	2.67 0.07
Explain things in a way that was easy to understand?			2.84	0.04	2.78	0.06	2.88 0.07	2.88 0.06
Seem to know the important information about the patient's medical history?			2.98	0.03	2.99	0.06	2.95 0.07	2.98 0.06
CAHPS Global Rating								
Using any number from 0 to 10, where 0 is the worst doctor possible and 10 is the best doctor possible, what number would you use to rate this doctor?			4.51	0.10	4.37	0.17	4.62 0.19	4.56 0.17

*Mean responses did not differ by race/ethnicity ($p > .05$ for all comparisons of African American and Latino to White).

[†]Responses are 1 = not at all; 2 = very little; 3 = to some extent; 4 = to a great extent.

[‡]1 = lowest physician responsiveness; 5 = highest physician responsiveness.

used for responses to the vignettes), which attempted to maximize power to detect racial/ethnic differences by pooling across outcomes, also failed to find significant evidence of differences ($p > .05$ in all instances).

In addition, while white, African American, and Latino respondents assigned similar adjusted mean 0–10 ratings (4.37, 4.62, and 4.56, respectively, $p > .05$) to the doctor in the video, the standard deviations for African Americans (2.63) and Latinos (2.59) were significantly greater than for whites (2.19, $p < .05$ in each case, by the Levene test). Similarly, African American and Latino respondents were more likely to use responses at both ends of the scale than white respondents. In particular, African Americans and Latinos were more likely than whites to use both the bottom two response options (14 and 12 percent versus 6 percent) and the top two response options (7 and 8 percent versus 3 percent); ORs = 1.90–2.81, $p < .05$ for all four contrasts from multinomial logistic regression. This reflects greater use of the extremes of the scales, or greater ERT, among African American and Latino respondents. This has also been observed in real-world CAHPS data, particularly for Latinos (Elliott et al. 2009b). There was no corresponding evidence of differences in ERT for the more specific report items (data not shown).

With respect to the index of perceived positive physician behaviors in the video, mean responses fell between 2 (very little) and 3 (to some extent). Mean perceptions of positive behaviors did not differ significantly by race/ethnicity (2.62 for African American, 2.56 for Latino, and 2.52 for white respondents; $p > .2$ for each comparison versus white, not shown). Mean responses for perceptions of negative physician behaviors occurred somewhat more often, nearly corresponding to a value of 3 (to some extent). The mean frequency of negative behaviors perceived by African Americans (2.73) was significantly lower than that for whites (2.93; $p = .01$); Latinos (2.83) did not significantly differ from whites ($p > .2$). These patterns are also consistent with respondents generally perceiving the video interaction as having at least as many negative as positive behaviors.

Finally, Table 4 includes the results of regressing the 0–10 CAHPS global rating of the physician on race/ethnicity, the positive and negative perception scales, and interactions between these. Because of the presence of interaction terms with race/ethnicity, the "positive behavior" and "negative behavior" coefficients estimate those coefficients within the reference group of non-Hispanic whites. As expected, perceptions of positive behavior were positively associated with the global ratings and perceptions of negative behaviors were negatively associated with this rating ($p < .0001$ for each). The magnitude of this coefficient was twice as large for positive perceptions as for

Table 4: Multivariate Regression Predicting Global Rating of Physician Quality from Race/Ethnicity and Positive and Negative Perceptions of Physician Behavior in the Video Interaction*

	Coefficient	Standard Error	p-Value
Intercept	− 2.30	0.51	<.0001
Race/ethnicity[†]			
African American	0.78	0.63	.21
Latino	0.40	0.61	.51
Mean (1–4) perceived behaviors[‡]			
Positive	1.88	0.24	<.0001
Negative	− 0.95	0.21	<.0001
Interaction terms			
African American × positive	− 0.23	0.33	.49
African American × negative	0.13	0.29	.64
Latino × positive	− 0.26	0.32	.42
Latino × negative	− 0.05	0.28	.86

*Case-mix-adjusted for age, education, and gender; coefficients not shown.
[†]White is the reference category.
[‡]1 = not at all, 2 = very little, 3 = to some extent, 4 = to a great extent.

negative perceptions, perhaps suggesting that the absence of positive perceptions may more strongly drive poor overall assessments of physicians than the presence of negative perceptions. Both the main effects of race/ethnicity and the interaction terms were nonsignificant, which is consistent with perceptions of physician behavior having a similar influence on 0–10 ratings of physicians across racial/ethnic groups. The nonsignificant interactions also suggest that the larger role of positive than negative perceptions is consistent across racial/ethnic groups.

DISCUSSION

We describe an experimental approach using standardized written and video vignette medical encounters to learn how differences in expectations, perceptions, and scale use may explain observed racial/ethnic differences in reported patient experiences in the real world. As has been demonstrated in other substantive areas (King et al. 2004), a vignette-based approach proved helpful for studying the extent to which racial/ethnic groups use response scales differently for assessing similar experiences; this approach may be

useful in further research that examines underlying causes of racial/ethnic disparities in care.

Our results suggest that different racial/ethnic groups have generally similar expectations regarding physicians' behaviors, with the exception of the extent to which they treat all patients fairly regardless of race. This indicates that the differences previously observed in real-world CAHPS data are not likely to be caused by some racial/ethnic groups having lower expectations of physicians, and therefore being more easily satisfied.

Using written vignettes that depicted a range of physician responsiveness to patient concerns, we find no evidence that CAHPS Clinician and Group report items are used differently by African Americans, Latinos, and whites in response to the same encounter stimuli. To increase the verisimilitude of the depicted encounter, we subsequently employed a single video portraying poorer-than-average physician behavior. In response to this video, all racial/ethnic groups provided similar responses on CAHPS Clinician and Group report items. These groups also had similar mean responses on 0–10 global ratings of the physician, and these ratings were similarly responsive to perceptions of positive and negative physician behavior. At the same time, however, African American respondents perceived fewer negative behaviors on the part of the physician than white and Latino respondents. This suggests that members of different racial/ethnic groups may explain identical observed behaviors in varying ways or may differentially value some behaviors and may warrant further exploration.

In the present study, Latinos and African Americans used both the positive and negative extremes of the response scale more often than whites for a standardized encounter. This extends previous findings of greater ERT by Latinos relative to whites in real-world encounters (Weech-Maldonado et al. 2008; Elliott et al. 2009b). Here, this tendency did *not* result in higher mean 0–10 ratings for Latinos and African Americans than for whites because the portrayal of poorer-than-typical physician behavior resulted in an atypical symmetric distribution of ratings (with ERT boosting both ends), rather than the more typical asymmetric distribution of ratings around a mean of 8 or 9 (with ERT boosting the positive end more than the negative and shifting the mean upward). Perhaps as a result, we find similar mean 0–10 ratings across racial/ethnic groups in response to this particular simulated encounter, unlike Elliott et al. (2009b) and Weech-Maldonado et al. (2008). These ERT differences might have resulted in higher mean 0–10 ratings for Latinos and African Americans if more positive physician behavior had been portrayed in the video encounter. Taken together, these findings reinforce the interpretation

that previously observed differences in 0–10 global ratings are most likely an artifact of scale use rather than an indication that Latinos and others with high ERT have more variable health care experiences. While it remains possible that minority patients are more likely than whites to have both the very best and worst health care experiences, the present findings cast doubt on this possibility.

This study has several limitations. First, despite efforts to achieve a representative sample and a demographic profile of respondents that reflects national distributions, respondents who agree to participate in an ongoing online panel may differ in unmeasured ways from the general public; our sample also differed in minor known ways from the full Knowledge Networks panel and the U.S. population as a whole. Second, our study was administered only in English, limiting our ability to comment on Latino patients who respond to CAHPS in Spanish, whose ERT and other differences from non-Latino whites are particularly large (Weech-Maldonado et al. 2008). Third, due to sample size limitations, we were unable to study Asians, a group for whom some of the strongest differences in CAHPS information have been observed (Weech-Maldonado et al. 2001, 2004). Fourth, the present study was administered via the Internet, rather than the mail and telephone modes currently recommended for the CAHPS Clinician and Group Survey (Agency for Healthcare Research and Quality 2008). However, a mode experiment using similar items from the CAHPS Hospital Survey found that responses to CAHPS items are quite similar in mail and Internet modes (Elliott et al. 2009a). Fifth, the current study examined only evaluations of physicians; future research might address the extent to which differences in expectations play a larger role in racial/ethnic differences in assessments of access to care or customer service. Finally, our focus here was limited to investigating the extent to which previous findings of differences in mean patient experience evaluations and ERT observed for patients' real-world encounters might also occur in response to standardized scenarios. Such an approach does not establish full measurement equivalence for these measures across racial/ethnic groups. Larger scale item response theory work on responses to vignettes such as these could substantially extend our findings; nonetheless, the more limited scope here provides insight into the interpretation of previously observed differences.

In addition, it is important to note that respondents may use modified CAHPS items and response scales that were adapted to the experiences of a third party in a single encounter (rather than for themselves over a period of time) somewhat differently than they use CAHPS measures of experiences for

real-world settings regarding their own care. This points to a challenge in accurately assessing experiences as complex as clinical encounters: standardized scenarios lack a sense of realism and personal meaning, but inquiring only about real-world encounters confounds the effects of respondent characteristics on response patterns with the content of clinical encounters.

Our findings suggest that African American, Latino, and white respondents have similar perceptions of the quality of physician–patient interactions when presented with the same behaviors, and are likely using CAHPS report items and composites similarly. With respect to the 0–10 global ratings, the finding of differential use of the extremes of the scales provides further evidence against the use of these items to assess racial/ethnic disparities. This finding is consistent with broader findings that 0–10 CAHPS global rating scales are more sensitive to the characteristics of respondents than more specific and concrete items—such as those related to physician communication—comprising the CAHPS report composites (e.g., Elliott et al. 2008).

Future work using multiple videos to consider a broader range of physician behaviors, perhaps manipulated over multiple dimensions such as patient-centeredness or patient activation (Hibbard et al. 2004), would enrich the current findings. Nonetheless, these findings provide a basis for two specific sets of actions. First, implementation of the 2008 MIPPA should emphasize CAHPS reports and composites rather than 0–10 global ratings in congressionally mandated public reporting of beneficiary experience by race/ethnicity. Second, our evidence supplies researchers and policy makers with a stronger basis for interpreting differences in CAHPS report and composite items as reflecting true disparities in need of remedy, and it supports continued efforts to address these problems.

ACKNOWLEDGMENTS

REFERENCES

Agency for Healthcare Research and Quality. 2008. "CAHPS Clinician & Group Survey and Reporting Kit 2008: Fielding the CAHPS Clinician & Group Survey" [accessed November 15, 2010]. Available at https://www.cahps.ahrq.gov/CAHPSkit/files/33_CG_Field ing_the_Survey.pdf

Baker, L., T. H. Wagner, S. Singer, and M. K. Bundorf. 2003. "Use of the Internet and E-Mail for Health Care Information: Results from a National Survey." *Journal of the American Medical Association* 289 (18): 2400–6. Appendix: "Validity of the Survey of Health and Internet and Knowledge Network's Panel and Sampling" [accessed on November 15, 2010]. Available at http://www.cdc.gov/pcd/issues/2004/oct/pdf/04_0004_01.pdf

Burke, B. L., H. Arkowitz, and M. Menchola. 2003. "The Efficacy of Motivational Interviewing: A Meta-Analysis of Controlled Clinical Trials." *Journal of Consulting and Clinical Psychology* 71: 843–61.

Dayton, E., C. Zhan, J. Sangl, C. Darby, and E. Moy. 2006. "Racial and Ethnic Differences in Patient Assessments of Interactions with Providers: Disparities or Measurement Biases?" *American Journal of Medical Quality* 21 (2): 109–14.

DeNavas-Walt, C., B. D. Proctor, and J. C. Smith. 2008. *Income, Poverty, and Health Insurance Coverage in the United States: 2007. Current Population Reports P60-235, U.S. Census Bureau*. Washington, DC: U.S. Government Printing Office.

Elliott, M. N., M. K. Beckett, K. Chong, K. Hambarsoomians, and R. D. Hays. 2008. "How Do Proxy Responses and Proxy-Assisted Responses Differ from What Medicare Beneficiaries Might Have Reported about Their Health Care?" *Health Services Research* 43 (3): 833–48.

Elliott, M. N., J. Brown, E. Goldstein, W. Lehrman, K. Hambarsoomian, and L. Giordano. 2009a. *Findings from the Second HCAHPS Mode Experiment*. Presented by M. N. Elliott at the Academy Health Annual Research Meeting, Chicago, IL, June 2009 and by J. Brown at the American Association for Public Opinion Research Annual Meeting, Hollywood, FL, May 2009.

Elliott, M. N., A. M. Haviland, D. E. Kanouse, K. Hambarsoomian, and R. D. Hays. 2009b. "Adjusting for Subgroup Differences in Extreme Response Tendency in Ratings of Health Care: Impact on Disparity Estimates." *Health Services Research* 44 (2, Part 1): 542–61.

Gallagher, P., E. J. Fowler, and P. Cleary. 2004. *Using Anchoring Vignettes to Assess Cross Cultural Comparability of CAHPS Ratings: A Pilot Study*. Presented at the 9th Annual CAHPS User Group Meeting, Baltimore, MD, December 2004.

Gary, L. C. 2006. *To Voice or To Exit: Racial Differences in Consumer Responses to Problematic Health Care Experiences*. Ph.D. dissertation, Yale University.

Hibbard, J. H., J. Stockard, E. R. Mahoney, and M. Tusler. 2004. "Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patients and Consumers." *Health Services Research* 39 (4, Part 1): 1005–26.

Jackson, J. L., J. Chamberlin, and K. Kroenke. 2001. "Predictors of Patient Satisfaction." *Social Science and Medicine* 52 (4): 609–20.

King, G., C. J. L. Murray, J. A. Salomon, and A. Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 97 (4): 567–83.

Knowledge Networks. No Date a. "KnowledgePanel Design Summary" [accessed on November 15, 2010]. Available at http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel%28R%29-Design-Summary-Description.pdf

Knowledge Networks. No Date b. *Description of KN Capabilities.* Personal Communication, William McCready, Knowledge Networks, March 2008.

Lurie, N., C. Zhan, J. Sangl, A. S. Bierman, and E. S. Sekscenski. 2003. "Variation in Racial and Ethnic Differences in Consumer Assessments of Health Care." *American Journal of Managed Care* 9 (7): 502–9.

Miller, W. R., and S. Rollnick. 1991. *Motivational Interviewing: Preparing People to Change Addictive Behavior.* New York: Guilford Press.

Morales, L. S., M. N. Elliott, R. Weech-Maldonado, K. L. Spritzer, and R. D. Hays. 2001. "Differences in CAHPS Adult Survey Reports and Ratings by Race and Ethnicity: An Analysis of the National CAHPS Benchmarking Data 1.0." *Health Services Research* 36 (3): 595–617.

Noble, P. C., M. A. Conditt, K. F. Cook, and K. B. Mathis. 2006. "The John Insall Award: Patient Expectations Affect Satisfaction with Total Knee Arthroplasty." *Clinical Orthopaedics and Related Research* 452: 35–43.

Schlenger, W. E., J. M. Caddell, L. Ebert, B. K. Jordan, K. M. Rourke, D. Wilson, L. Thalji, J. M. Dennis, J. A. Fairbank, and R. A. Kulka. 2002. "Psychological Reactions to Terrorist Attacks: Findings from the National Study of Americans' Reactions to September 11." *Journal of the American Medical Association* 288 (5): 581–8.

Silver, R. C., E. A. Holman, D. N. McIntosh, M. Poulin, and V. Gil-Rivas. 2002. "Nationwide Longitudinal Study of Psychological Responses to September 11." *Journal of the American Medical Association* 288 (10): 1235–44.

Uhrig, J., J. S. Laufenberg, C. F. Eicheldinger, S. Bernard, and G. G. Brown. 2004. *How Do Medicare Beneficiaries' Assessments of Care Differ by Race and Ethnicity?* Presentation at the 2004 Annual Meeting of the American Public Health Association, Washington, DC.

U.S. Census Bureau. 2009a. "The 2009 Statistical Abstract. Table 7. Resident Population by Age and Sex: 1980 to 2007" [accessed on November 15, 2010]. Available at http://www.census.gov/prod/2008pubs/09statab/pop.pdf

U.S. Census Bureau. 2009b. "The 2009 Statistical Abstract. Table 23. Population by Core-Based Statistical Area Status and State: 2000 to 2007" [accessed on November 15, 2010]. Available at http://www.census.gov/prod/2008pubs/09statab/pop.pdf

U.S. Census Bureau. No Date a. "Fact Sheet, United States: 2006–2008 American Community Survey 3-Year Estimates Data Profile Highlights" [accessed on November 15, 2010]. Available at http://factfinder.census.gov/servlet/ACSSAFFFacts?_submenuId=factsheet_0 &_sse=on

U.S. Census Bureau. No Date b. "American FactFinder. United States – Region, Division, and States: GCT-T1 2008 Population Estimates" [accessed on November 15, 2010]. Available at http://factfinder.census.gov/servlet/GCTTable?_bm=y

&-context=gct&-ds_name=PEP_2008_EST&-mt_name=PEP_2008_EST_GCTT1_
US41&-CONTEXT=gct&-tree_id=808&-redoLog=false&-geo_id=&-format=US-
41&-_lang=en

Volandes, A. E., L. S. Lehmann, E. F. Cook, S. Shaykevich, E. D. Abbo, and M. R.
Gillick. 2007. "Using Video Images of Dementia in Advance Care Planning."
Archives of Internal Medicine 167 (8): 828–33.

Wagner, T. H., L. C. Baker, M. K. Bundorf, and S. Singer. 2004. "Use of the Internet for
Health Information by the Chronically Ill." *Preventing Chronic Disease* 1 (4): A13.

Weech-Maldonado, R., M. N. Elliott, L. S. Morales, K. Spritzer, G. N. Marshall, and
R. D. Hays. 2004. "Health Plan Effects on Patient Assessments of Medicaid
Managed Care among Racial/Ethnic Minorities." *Journal of General Internal
Medicine* 19 (2): 136–45.

Weech-Maldonado, R., M. N. Elliott, A. Oluwole, K. C. Schiller, and R. D. Hays. 2008.
"Survey Response Style and Differential Use of CAHPS Rating Scales by
Hispanics." *Medical Care* 46 (9): 963–8.

Weech-Maldonado, R., L. S. Morales, M. N. Elliott, K. L. Spritzer, G. Marshall, and
R. D. Hays. 2003. "Race/Ethnicity, Language and Patients' Assessments of Care
in Medicaid Managed Care." *Health Services Research* 38 (3): 789–808.

Weech-Maldonado, R., L. S. Morales, K. Spritzer, M. Elliott, and R. D. Hays. 2001.
"Racial and Ethnic Differences in Parents' Assessments of Pediatric Care in
Medicaid Managed Care." *Health Services Research* 36 (3): 575–94.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this
article:

> Appendix SA1: Author Matrix.
> Appendix SA2: Vignettes.
> Appendix SA3: Transcription of Video.
> Appendix SA4: Examining Perceptions of Physician Behavior.

Please note: Wiley-Blackwell is not responsible for the content or func-
tionality of any supporting materials supplied by the authors. Any queries
(other than missing material) should be directed to the corresponding author
for the article.